

Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images

Stefanus Kieu Tao Hwa¹, Mohd Hanafi Ahmad Hijazi², Abdullah Bade³, Razali Yaakob⁴, Mohammad Saffree Jeffree⁵

^{1,3}Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Malaysia

²Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia

⁴Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

⁵Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Malaysia

Article Info

Article history:

Received Aug 13, 2019

Revised Nov 5, 2019

Accepted Nov 20, 2019

Keywords:

Canny edge detector

Deep learning

Ensemble

Medical image analysis

Tuberculosis detection

ABSTRACT

Tuberculosis (TB) is a disease caused by Mycobacterium Tuberculosis. Detection of TB at an early stage reduces mortality. Early stage TB is usually diagnosed using chest x-ray inspection. Since TB and lung cancer mimic each other, it is a challenge for the radiologist to avoid misdiagnosis. This paper presents an ensemble deep learning for TB detection using chest x-ray and Canny edge detected images. This method introduces a new type of feature for the TB detection classifiers, thereby increasing the diversity of errors of the base classifiers. The first set of features were extracted from the original x-ray images, while the second set of features were extracted from the edge detected image. To evaluate the proposed approach, two publicly available datasets were used. The results show that the proposed ensemble method produced the best accuracy of 89.77%, sensitivity of 90.91% and specificity of 88.64%. This indicates that using different types of features extracted from different types of images can improve the detection rate.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Mohd Hanafi Ahmad Hijazi,
Faculty of Computing and Informatics,
Universiti Malaysia Sabah,
Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
Email: hanafi@ums.edu.my

1. INTRODUCTION

Tuberculosis (TB) is one of the deadliest infectious diseases in the world. A bacteria called Mycobacterium Tuberculosis causes it. TB not only attack the lungs but can also infect other body parts like bones or spine. Tuberculosis is spread through the air, usually from coughs and sneezes by people with active tuberculosis disease. According to the World Health Organization, TB is one of the top 10 causes of death worldwide [1]. In 2017, 10 million people fell ill with TB, and 1.6 million died from the disease. In Malaysia, the approximately 25173 cases of TB are reported in 2018. X-rays can detect TB to a certain extent, but it cannot guarantee whether the patient has TB or some other infection. TB is a curable disease. Hence, early detection of TB is critical to increasing the chances of recovery [2]. However, in Malaysia, the late diagnose of TB, in most cases, reducing the chances of survival [1]. For centuries, radiologist faced the challenge of differentiating tuberculosis and lung cancer because they mimicked each other [3]. The lack of publically available datasets also makes it difficult to provide more computer-aided detection systems [4]. There is also a lack of radiology interpretation expertise in TB prevalent places [5]. A system to semi-automatically classifying pulmonary nodules with low false positivity is thus considered necessary to help radiologists screens the chest x-ray images [6], as medical diagnosis is one of the most important issues in healthcare [7]. Image processing has shown potential in tuberculosis detection. Recently, deep learning has been applied to

analyze features from medical images [8]. Deep learning automatically generates hierarchical features from images. The features at the higher level are generated from the features at the lower level. It was shown that the deep learning ability to learn features at the higher level produced better classification results [9]. Works on the usage of deep learning on chest x-rays images to detect TB have been done before. In [10], multi-level image enhancement were performed on the x-ray lung images. Then, a backpropagation neural network was used to classify TB. A completely automatic frontal chest radiograph screening system able to detect lungs infected with TB was presented [11]. This method begins with an atlas-based lung segmentation algorithm, then it extracts manually selected features such as shape and curvature descriptor histograms or the eigenvalues of the hessian matrix. In the end, a classifier is used to diagnose the disease. In [12], a simple CNN is used to perform the same task. Another CNN-based model to classify different categories of TB manifestations was developed in [13]. First, region proposals were extracted. Then, global and local features were extracted. Afterward, for each region, a CNN model is trained to calculate the new features for further classification. Finally, the Support Vector Machine is applied for final region classification and TB manifestations recognition. A deep learning-based automatic detection (DLAD) algorithm was developed for TB classification in [14]. The deep CNN used in this DLAD algorithm is comprised of 27 layers with 12 residual connections. Six datasets were used for training and testing. DLAD shows consistent performance in the detection of TB on chest x-rays, outperforming physicians and thoracic radiologists. A technique which included demographic information to improve the CNN's performance was introduced [15]. Age, weight, height, and gender were listed as demographic variables. Results show that CNN including demographic variable has a higher AUC score and greater sensitivity than CNN based on chest x-rays images only. [16] Performed TB detection using transfer learning from ImageNet and training on a dataset of 10848 chest x-rays. In [17], pre-training was done on the NIH-14 dataset, and then the features learned from the NIH dataset is transferred to TB datasets. Experiments show that the features transferred is useful for identifying TB.

When more than one model is used to make a prediction, this is known as ensemble learning. Ensemble reduces the variance of predictions, thus providing predictions that are more accurate than any single model. An ensemble created by feature-level fusion of three deep neural network models was also used to classify TB [18]. These three models are ResNet, Inception-ResNet and DenseNet, thus the ensemble was named as RID network. The models were used as feature extractors and SVM was used as a classifier. TB classification was also done using another ensemble of three standard architectures, namely AlexNet, GoogleNet and ResNet [19]. Each architecture was trained from scratch, and they used different optimal hyper-parameter values. The accuracy, sensitivity and specificity of the ensemble are higher than when each of the standard architecture was used individually. Fine-tuned AlexNet, VGG-16, VGG-19, ResNet-50, ResNet-101 and ResNet-512 were used by [20] to classify TB. An ensemble of these six CNNs was built. The ensemble models were obtained by using simple linear averaging of the probability predictions given by the individual models. Pre-trained AlexNet and GoogLeNet were used to perform pulmonary TB classification in [21], and they found that higher accuracy was obtained when using the pre-trained model. Later, they ensembled these models by using the weighted averages of each model's probability scores. The work presented in [22] used pre-trained CNN classifiers for TB detection, where majority voting was employed to ensemble the generated classifiers. Lopes and Valiati proposed a Bag of CNN features to classify TB [23] where GoogLeNet, VggNet, and ResNet are used to extract features. Then, each chest radiograph is divided into subregions, whose size is equal to the network's input layer. Each subregion is an instance, or "feature", and each radiograph is a "bag". They then created a Bag of Features Ensemble by using a simple soft-voting scheme.

Based on the literature, deep learning has produced good results for TB detection. All of the works described above extract features from the original x-ray images of the chest. However, most studies used features that were automatically extracted by CNN. Only a handful of studies focused on non-CNN-extracted features. Different features should be explored because the features used influences the performance of the classifier. For an ensemble of classifiers to perform well, it requires a diversity of errors [24]. In other words, the errors of the base classifiers should have a low correlation. Most of the studies only combine classifiers that were trained on similar features. This paper presents an ensemble deep learning for TB detection using chest x-ray and Canny edge detected images. This method introduces a new type of feature for the TB detection classifiers, thereby increasing the diversity of errors of the base classifiers. We conjectured that using different type of images and ensemble classifiers will produce a better TB detection rate. This paper has two contributions as follows:

- Generation of CNN classifiers based on two types of images, which are the original chest-x-rays and the Canny edge detected chest x-rays, for TB detection.
- Ensemble technique that uses average probability scores to combine the different classifiers employed in TB detection.

The proposed approach that employs ensemble deep learning and Canny edge detected images to detect TB is presented in Section 2. The experimental setup and discussion of results are described in Section 3. The conclusion is presented in Section 4.

2. THE PROPOSED APPROACH

In this paper, an approach that employs ensemble deep learning coupled with Canny edge detector is proposed. It consists of three phases: (i) image pre-processing, (ii) classifier generation, and (iii) ensemble classification. Figure 1 shows the image pre-processing and classifier generation phases. The ensemble classification phase is shown in Figure 2. Sub-sections 2.1 to 2.3 describes the details of each phase.

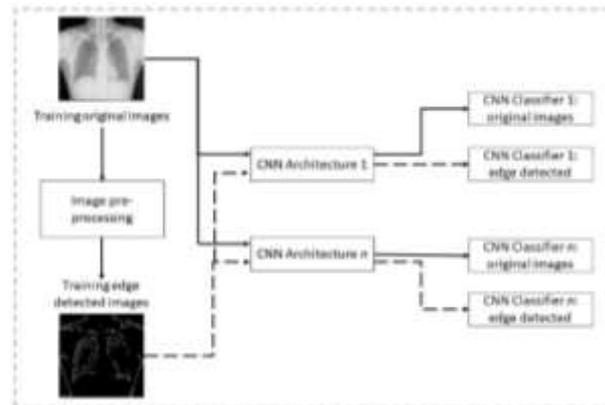


Figure 1. Image pre-processing and classifier generation modules

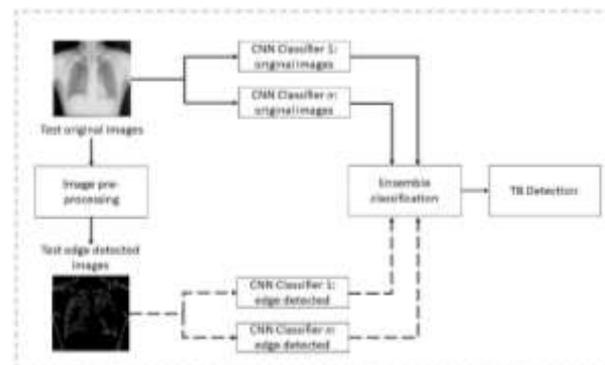


Figure 2. Ensemble classification module

2.1. Image pre-processing

The image acquired were first pre-processed to obtain an alternate form of feature, namely edge feature. The pre-processing of the images has two stages. The first stage is the image resizing. All images were resized to 250 x 250 pixels. This is to ensure all image sizes are uniform. It is also done so that the image size matches the input size of the CNNs. The second stage is Canny Edge detection. In this step, the images are processed such that they only contain the edges. The idea was that images with TB may have more unusual edges that the normal images which could increase the detection rate of TB. Figure 3 shows an example of an x-ray image before the pre-processing, and Figure 4 shows the image after the pre-processing.



Figure 3. Chest x-ray image before pre-processing

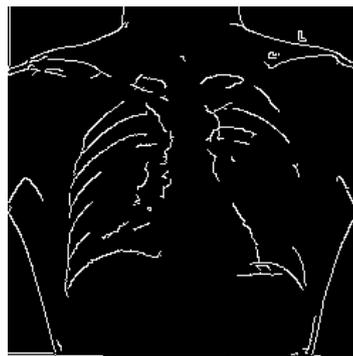


Figure 4. Chest x-ray image after pre-processing

2.2. CNN classifier generation

In this stage, various CNN architectures could be employed to generate the classifiers. For the work presented in this paper, more than one CNN architecture should be used. Each of the selected CNN architecture will learn features from the original image and the edge detected image individually. For each CNN architecture, two classifiers will be generated from the original image and the edge detected image.

2.3. Ensemble classification

In this phase, the classification commences with the detection of TB by an individual CNN classifier. Ensemble classification was then conducted to produce better detection results. To achieve this, the average scoring mechanism was employed. To predict the final label of each test image (TB or non-TB), the prediction score generated by the individual CNN classifier was averaged, and the resulting probability score determines the final label.

2.4. Performance measure

To measure the performance of the proposed approach, three metrics were employed: the sensitivity, specificity and accuracy. Sensitivity is used to measure the ability of the model to identify positive cases, while specificity measures how well the model to identify negative cases. The overall performance of the model is indicated by the accuracy. In our case, a positive case representing TB, while negative case representing non-TB.

3. EXPERIMENTAL SETUP AND DISCUSSION OF RESULTS

This section describes the dataset used to evaluate the proposed work, the experiments conducted, and the discussions of results.

3.1. The dataset

To evaluate the work presented in this paper, two public chest x-ray images datasets are used, which are Montgomery and Shenzhen datasets [25]. The Montgomery dataset has a total of 138 images, which consists of 80 normal lung images and 58 images of TB. The size of the chest x-ray is either 4020×4892 pixels or 4892×4020 pixels in Portable Network Graphics (PNG) format. The Shenzhen dataset has a total of 662 images, which consists of 326 normal lung images and 336 TB images. The size of the images is approximately 3000×3000 pixels, also in PNG format. The total number of healthy lung images is 406, while the total number of TB infected lung images is 394. In order to achieve a balanced dataset, 12 normal images were randomly omitted, such that there is an equal number of samples in each class. From the selected images, 90% of them were used for training and 10% for testing. The train-test split ratio selected is similar to the work presented in [26] and [18]. Ten-cross validation (TCV) was used to validate the training data.

3.2. Experimental setup

To measure the performance of the proposed approach, two sets of experiments were conducted. The first was to evaluate the performance of features extracted from the original and edge detected images for TB detection. These features were used to generate CNN based classifiers, whereby these classifiers were then used to detect TB cases. The second experiment was conducted to compare the performance of the ensemble of classifiers used in experiment 1 with individual classifiers. Concerning the work presented in

this paper, the Keras implementation of InceptionV3 and VGG16 were used. Both architectures were pre-trained using the ImageNet dataset.

3.3. Experimental results

The results of the sets of the experiment described above are presented in Sub-section 3.3.1 and 3.3.2 respectively.

3.3.1. Performance of different set of images and CNN architectures to detect TB

In this set of experiment, the images were classified as either TB or non-TB. For each set of images and CNN architecture, the learning rate and epoch were set to 0.0001 and 2000, respectively. The average validation results, produced by the training data, are shown in Table 1. TCV was used to validate the generated classifiers. The best detection rate was obtained when applying VGG16 on the original x-ray images. The generated classifiers were subsequently tested using the test data. Table 2 shows the results of the experiments. VGG16 obtained the best sensitivity on the original dataset and also specificity on the edge dataset. VGG16 on both sets of images recorded the same accuracy. Better detection results were recorded on the test data.

Table 1. Validation result of CNN Architectures to TB Detection on original and edge dataset

| Image type | CNN Architecture | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---------------|------------------|-----------------|-----------------|--------------|
| Original | VGG16 | 85.43 | 90.00 | 87.71 |
| | InceptionV3 | 83.14 | 84.00 | 83.57 |
| Edge detected | VGG16 | 80.57 | 84.86 | 82.72 |
| | InceptionV3 | 76.86 | 81.43 | 79.14 |

Table 2. Performance of Different CNN Architectures to TB Detection on original and edge dataset

| Image type | CNN Architecture | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---------------|------------------|-----------------|-----------------|--------------|
| Original | VGG16 | 86.36 | 90.91 | 88.64 |
| | InceptionV3 | 77.27 | 81.82 | 79.55 |
| Edge detected | VGG16 | 84.09 | 93.18 | 88.64 |
| | InceptionV3 | 77.27 | 75.00 | 76.14 |

3.3.2. Performance of ensemble classifiers to TB detection

For the second set of experiments, the individual CNN classifiers were ensemble to perform the classification. All the possible combinations were recorded. Table 3 shows the results obtained when applied to the test data. (A denotes VGG16 trained on the original dataset, B denotes VGG16 trained on edge dataset, C denotes InceptionV3 trained on the original dataset, D denotes InceptionV3 trained on edge dataset)

Table 3. The performance of ensemble CNN for TB detection

| CNN classifier | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|----------------|-----------------|-----------------|--------------|
| AB | 88.64 | 93.18 | 90.91 |
| AC | 86.36 | 86.36 | 86.36 |
| AD | 84.09 | 84.09 | 84.09 |
| BC | 84.09 | 86.36 | 85.23 |
| BD | 81.82 | 84.09 | 82.95 |
| CD | 81.82 | 84.09 | 82.95 |
| ABC | 88.64 | 95.45 | 92.05 |
| ABD | 84.09 | 90.91 | 87.50 |
| ACD | 86.36 | 86.36 | 86.36 |
| BCD | 79.55 | 86.36 | 82.95 |
| ABCD | 90.91 | 88.64 | 89.77 |

3.4. Analysis of results

Based on the results shown in Tables 1 and 2, it is observed that VGG16 performs the best when applied to the original x-ray images. The classifiers produced better results when applied to the test data as shown in Table 2. When applied individually, InceptionV3 does not perform well. It is observed that all the ensemble combination produces better accuracy than any individual InceptionV3 classifier. Also, there are three ensemble combination produced better accuracy than any individual VGG16 classifier. These three combinations are AB, ABC and ABCD. The ensemble combination that produces the highest accuracy is ABC, at 92.05%, with sensitivity and specificity of 88.64% and 95.45% respectively. However, the ensemble that produces the highest sensitivity is ABCD, at 90.91%, with specificity and accuracy of 88.64% and

89.77% respectively. Sensitivity is considered pertinent in medical image analysis as we want to reduce false negatives. The experiments also show that by using more than one type of images, a better detection rate can be achieved. Based on this result, it shows that using features automatically extracted from a different type of images produce better TB detection.

4. CONCLUSION

An approach to detect TB using an ensemble of CNN architecture on different type of images features, extracted from the original image and Canny edge detected image, is presented. Two different CNN architectures were employed, namely VGG 16 and InceptionV3. The results obtained shows that, the utilization of ensemble classifiers using averaged probability decision and variation of features used, produced better TB detection performance. This indicates that using different types of features extracted from different types of images can improve the detection rate. Future work may focus on other types of features to further improve the detection rate and reducing the processing time. We also would extend the scope to classify the TB based on the severity.

ACKNOWLEDGEMENTS

Universiti Malaysia Sabah supports this work through a grant SDK0018-2017 and Knowledge Technology Research Unit, UMS.

REFERENCES

- [1] World Health Organisation, *Global Health TB Report*. 2018.
- [2] A. Yahiaoui *et al.*, "A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines", *Biomed. Res.*, vol. 28, no. 9, pp. 4208–4212, 2017.
- [3] I. Hammen, "Tuberculosis mimicking lung cancer," *Respiratory Medicine Case Reports*, vol. 16, pp. 45–47, 2015.
- [4] J. Monsi *et al.*, "XRAY AI: Lung Disease Prediction Using Machine Learning," *Int. J. Inf. Syst. Comput. Sci.*, vol. 8, no. 2, pp. 51–54, 2019.
- [5] J. Melendez *et al.*, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," *Scientific Reports*, vol. 6, no. 1, 2016
- [6] K. C. Siang *et al.*, "A Review of Lung Cancer Research in Malaysia," *Med. J. Malaysia*, vol. 71, pp. 70–78, 2016.
- [7] N. Walia *et al.*, "Effective Analysis of Lung Infection using Fuzzy Rules", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 5, no. 2, pp.55-63, 2016
- [8] G. Litjens *et al.*, "A survey on deep learning in medical image analysis", *Medical Image Analysis*, vol. 42, pp. 60-88, 2017.
- [9] S. S. Meraj *et al.*, "Artificial Intelligence in Diagnosing Tuberculosis : A Review," *Int. Journal on Advanced Sci., Eng. and Inform. Technology*, vol. 9, no. 1, pp. 81-91, 2019.
- [10] C. V *et al.*, "Multi-level Image Enhancement for Pulmonary Tuberculosis Analysis," *Int. J. Sci. Appl. Inf. Technol.*, vol. 1, no. 4, pp. 102–106, 2012.
- [11] S. Vajda *et al.*, "Feature Selection for Automatic Tuberculosis Screening in Frontal Chest Radiographs", *Journal of Medical Systems*, vol. 42, no. 8, 2018.
- [12] F. Pasa *et al.*, "Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization", *Scientific Reports*, vol. 9, no. 1, 2019.
- [13] Y. Cao *et al.*, "Improving Tuberculosis Diagnostics Using Deep Learning and Mobile Health Technologies among Resource-Poor and Marginalized Communities," *2016 IEEE 1st Int. Conf. Connected Health: Applications, Systems and Eng. Technologies (CHASE)*, Washington, DC, 2016, pp. 274-281.
- [14] E. Hwang *et al.*, "Development and Validation of a Deep Learning–based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs", *Clinical Infectious Diseases*, vol. 69, no. 5, pp. 739-747, 2018.
- [15] S. Heo *et al.*, "Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data," *Int. J. Environ. Res. Public Health*, vol. 16, 2019.
- [16] S. Heo *et al.*, "Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annu.Workers' Health Examination Data", *Int. Journal of Environmental Research and Public Health*, vol. 16, no. 2, p. 250, 2019.
- [17] Q. H. Nguyen *et al.*, "Deep Learning Models for Tuberculosis Detection from Chest X-ray Images," *2019 26th Int. Conf. Telecommun. (ICT)*, Hanoi, Vietnam, 2019, pp. 381-385.
- [18] R. Rashid *et al.*, "Hybrid RID Network for Efficient Diagnosis of Tuberculosis from Chest X-rays," *2018 9th Cairo Int. Biomedical Eng. Conf. (CIBEC)*, Cairo, Egypt, 2018, pp. 167-170.
- [19] R. Hooda *et al.*, "Automated TB classification using ensemble of deep architectures", *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31515-31532, 2019.
- [20] M. T. Islam *et al.*, "Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks," arXiv e-prints, 2017.

- [21] P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks", *Radiology*, vol. 284, no. 2, pp. 574-582, 2017.
- [22] M.H.A. Hijazi *et al.*, "Ensemble Deep Learning for Tuberculosis Detection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 10, no. 1, pp. 401-407, 2018.
- [23] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135-143, 2017.
- [24] T. Dietterich, "Ensemble Methods in Machine Learning" in *Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2000, pp.1-15.
- [25] S. Jaeger *et al.*, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imaging Med. Surg.*, vol. 4, no. 6, pp. 475-477, 2014.
- [26] J. Melendez *et al.*, "A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays," in *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 179-192, 2015.

BIOGRAPHIES OF AUTHORS



Stefanus Kieu Tao Hwa is a student of Mathematics and Computer Graphics at the Faculty of Science and Natural Resources, Universiti Malaysia Sabah in Malaysia. He is currently a postgraduate student in UMS.



Mohd Hanafi Ahmad Hijazi is an Associate Professor of Computer Science at the Faculty of Computing and Informatics, Universiti Malaysia Sabah in Malaysia. His research work addresses the challenges in knowledge discovery and data mining to identify patterns for prediction on structured and/ or unstructured data; his particular application domains are medical image analysis and understanding and sentiment analysis on social media data. He has authored/ co-authored more than 40 journals/ book chapters and conference papers, most of which are indexed by Scopus and ISI Web of Science. He also served on the program and organizing committees of numerous national and international conferences. He is the leader of Data Technologies and Applications research group at the faculty.



Dr. Abdullah Bade born in 1977, received his BSc. in Computer Science in 2000, MSc. in Computer Science (Computer Graphics) in 2003 from Universiti Teknologi Malaysia and PhD in Industrial Computing in 2008 from National University of Malaysia. Currently, he is an Associate Prof. in Computer Graphics and Image Processing at Faculty of Science and Natural Resources, Universiti Malaysia Sabah. Dr. Abdullah is a professional member of ACM SIGGRAPH, member Malaysian Society of Mathematics and Science (PERSAMA) and serves to the government as Malaysia Qualification Agency (MQA) panel of assessor in the field of Computing and Multimedia since 2005. His research interest includes Computer Graphics and Image Processing algorithms and all aspects of image generation. He had published numerous articles in refereed journals, conference proceedings, books and technical papers.



Assoc. Prof. Razali Yaakob received the Bachelor Degree in Computer Science in 1996 and Master in Computer Science from in 1999, from Universiti Putra Malaysia, and PhD from University of Nottingham, United Kingdom in 2008. Currently, he is a lecturer at the Faculty of Computer Science and IT, Universiti Putra Malaysia. His research areas include artificial neural network, pattern recognition, and evolutionary computation in game playing. He is a member of the Intelligent Computing Group at the faculty.



Professor Dr Mohammad Saffree Jeffree is the Dean of Faculty of Medicine and Health Sciences, University Malaysia Sabah. He holds a Medical Degree from Universiti Malaysia Sarawak and Master of Community Medicine (Occupational Health) from National University of Malaysia. He also obtained other qualifications such as Epidemic Intelligent Program from Ministry of Health Malaysia, Occupational Health Doctor from Department of Occupational Safety and Health Malaysia (DOSH), and Certified Medical Impairment Assessor from National Institute of Occupational Safety and Health (NIOSH). Before joining UMS, previously he works with Ministry of Health Malaysia as a Medical Officer for 9 years and as Public Health Medicine Specialist for 6 years. He is actively involved in occupational health, community health, infectious diseases, non-communicable diseases and cancer research areas. He is currently a supervisor for more than 10 postgraduate students and had published several papers in high impact journals.